ImpactEd Evaluation

Bookmark Reading Charity

Bookmark Reading Charity

IMPACT REPORT

2023/24

Contents

List of figures

List of tables

Executive summary

Bookmark Reading Charity aims to improve children's literacy by promoting a reading for pleasure culture in primary schools, with focus on supporting children in the most disadvantaged communities. Each one-to-one reading programme comprised 12 sessions, although pupils did not always participate in all sessions. Pupils could participate in one or more of Bookmark's One-to-one Reading Programmes at the discretion of the school.

This report discusses findings of an evaluation of Bookmark's programmes[1] over the 2023/24 academic year. It investigates the impact of the programme on outcomes in different demographic groups, progress in reading in Bookmark pupils compared with similar non-Bookmark pupils, and the impact of dosage through the number of sessions per programme, as well as the number of programmes completed by pupils throughout the academic year.

Key findings

Bookmark pupils, on average, increased their standardised reading score by 5.8[2] over the 2023/24 academic year, indicating that pupils' reading ability increased more than expected. This change was statistically significant (p<0.001), meaning this change was unlikely to have been recorded only by chance.

1 Does the Bookmark programme have a differential impact on specific subgroups?

Participating pupils eligible for Pupil Premium and pupils with Special Educational Needs made better improvements in reading scores than their non-Pupil Premium and non-SEN peers over the course of the academic year. All of these changes were statistically significant.

2 Do Bookmark pupils make more progress in reading than similar non-Bookmark pupils?

Pupils most in need of help with their reading were selected to participate in Bookmark's programme. This is evidenced in Bookmark pupils having a lower baseline reading score, of 93.9, than similar non-Bookmark pupils, of 103.9.

Because participating and control pupils had such different baselines, the groups were not similar in academic ability and so it was **not possible to meaningfully answer Research Question 2 using the data collected** in this evaluation[3]. Furthermore, there was no statistically significant difference between the change in reading scores in Bookmark and similar non-Bookmark pupils (p=0.61). Further evaluation is advised to find a significant difference, although it was a promising to not observe non-Bookmark pupils improving statistically significantly more than Bookmark pupils, and that the change in reading scores in both groups was similar.

[1] References made through the remainder of the report to 'programmes' refers to Bookmark's One-to-one Reading Programme

[2] See *2.2. Outcome measure: standardised reading scores* for how to interpret an increase in this measure

[3] See *6.2. Recommendations* for guidance on addressing this in future evaluations

3A Does the number of sessions in a programme impact the amount of reading progress a pupil makes?

Pupils who had participated in 6-12 sessions recorded an average increase in reading score, while those who completed 1-5 sessions in a single programme recorded average changes that were less consistent, with those participating in one, four and five sessions recording decreases. However, there was **no clear linear correlation between the number of sessions in a programme and the amount of progress a pupil made in standardised reading score** from before to after a programme. This indicates that linearly, participating in more programmes does not mean a pupil would make more progress in reading.

3B Does the number of programmes a pupil participates in impact the amount of reading progress they make?

Bookmark pupils who completed at least one programme increased their standardised reading scores more than those who did not complete any. Those who completed none recorded a very small increase (+0.2) while those who completed more recorded increases of 4.4 to 12.1. However, there were low sample sizes and varying levels of statistical significance here, so this analysis should be explored further, but the finding itself is promising. Moreover, there were no statistically significant linear correlations found between the number of programmes completed and change in reading.

Recommendations

More information on all the below future recommendations can be found in section *6.2. Recommendations.*

Key recommendations for any **future iterations of the programme** include:

- **Encourage pupils to complete a minimum of six sessions.** This reflects on findings from Research Question 3A.
- **Continue to not limit the number of programmes a pupil can do over the year**. This reflects on findings from Research Question 3B.
- **Consider tailoring the programme for different year groups more**. This reflects on findings from Research Question 1.
- **Continue targeting Pupil Premium pupils and those with Special Educational Needs**. This reflects on findings from Research Question 1.

Key recommendations for the **future evaluation design** include:

- **Review the methodology for Research Questions 2 and 3A to allow a longer window to observe impact.** This would help to understand whether the programme has any longer-term impact on pupil outcomes, and the data can be analysed more robustly.
- **Conduct subgroup analysis with similar participating and control pupils.** This report has determined there is differential impact on specific pupils, and using a control group would be beneficial in reinforcing these findings. It would also enable the exploration of why

some Bookmark pupils made better progress in reading than similar non-Bookmark pupils, while others did not.

- **Run a retrospective evaluation that matches participating pupils with similar pupils from newly partnered Bookmark schools, using baseline standardised reading scores as covariates in propensity score matching.** This would enable a comparison between pupils who are academically similar, as well as similar in demographic backgrounds, which would hopefully lead to a meaningful finding for Research Question 2. It would also allow further exploration into the finding from Research Question 3B regarding pupils who completed at least one programme improving more than pupils who did not complete any, but were meant to.
- **Re-investigate Research Question 2 (the amount Bookmark pupils progress compared to non-Bookmark pupils) using a methodology that looks at progress over the entire academic year**. Research Question 3B found a lack of a linear relationship between the number of programmes participated in and the change in reading score, indicating this dosage does not need to be accounted for – something that was done in this evaluation. So, it is fine to allow more variation in the amount of intervention a pupil participated inin order to allow more time for change to occur, making statistically significant findings would be more likely.
- **Analyse and interpret stakeholder voice alongside attainment data**, such as qualitative data from pupils, teachers or Bookmark reading volunteers**.** This could help to build a holistic understanding of impact.
- **Ask teachers to select an appropriate ethnic group for pupils, with the added option of 'Other (please state)'**, rather than allowing open text by default. While open text was opted for in this evaluation to address the lack of standard ethnicity classifications in schools, this made it more challenging than anticipated to aggregate this data. Changing to a selectable format would mean the number of ethnic subgroups is more limited, similar groups can be aggregated together and, as a result, sample sizes are larger. This would allow for more in-depth analysis between different ethnic groups to understand whether the programme has a differential impact.

1. Introduction

Bookmark Reading Charity aims to improve children's literacy by promoting a reading for pleasure culture in primary schools, with focus on supporting children in the most disadvantaged communities. This work is delivered through 10 literacy programmes. Their reading programme involves teacher-selected pupils in years 1-5 who are struggling to read. In each programme pupils engage in up to 12 sessions with a reading volunteer. In the sessions, Bookmark volunteers spend time reading stories and playing games with the pupils on a secure online platform or face-to-face to promote a joy of reading.

This evaluation assesses the impact of Bookmark's reading programme on attainment. For this evaluation, data from 265 Bookmark pupils (participating) and 648 non-Bookmark pupils (control) in years 2-5 from 18 partner schools was collected and analysed. This included **pupil demographic data such as year group, gender, Pupil Premium status, and standardised reading scores**.

There were four main windows of data collection:

- Baseline: end of 2022/23 academic year/ start of 2023/24 academic year
- Autumn term 2023
- Spring term 2024
- Summer term 2024

The analysis methodologies deployed were dependent on the individual research questions being investigated, though they always used a **systematic approach by using the data to quantitatively assess the outcomes and effectiveness** of the programme.

This report presents the findings of the evaluation. It discusses four key areas: an overview of the **progress made by participating pupils throughout the 2023/24 academic year**; the **differences between participating and control pupils**; the **impact that the programme had on different demographic groups**; the **impact of dosage** through the number of sessions per programme, and number of programmes participated in throughout the academic year.

2. Methodology

2.1. Research questions

As well as understanding the overall progress made by participating Bookmark pupils throughout the 2023/24 academic year, this report explores four key research questions, as listed below.

1 Does the Bookmark programme have a differential impact on specific subgroups?

2 Do Bookmark pupils make more progress in reading than similar non-Bookmark pupils?

3A Does the number of sessions in a programme impact the amount of reading progress a pupil makes?

3B Does the number of programmes a pupil participates in impact the amount of reading progress they make?

2.2. Outcome measure: standardised reading scores

Outcome	Quantitative measures
Change in reading	▶ Standardised reading scores o Scale 60-140, where 100 is the set national average. o Schools used various tests such as NFER, STAR and PiRA

Table 1: Outcome measures

Standardised reading scores reflect how close a pupil is working at the level they are expected to for their age in reading, where a score of 100 represents where a pupil is expected to be. A score lower than 100 means a pupil has not met the expected standard in the test, while a score higher than 100 means they are working above the expected standard.

When a standardised reading score is calculated, it considers a pupil's age. Thus, we are able to aggregate these scores across year groups in analysis. Furthermore, it means that just because a pupil has a higher standardised reading score than another pupil, this does not necessarily mean that they are therefore working at a higher level. It does, however, mean they are working higher relative to where they should be for their age (i.e. closer to the expected if below 100, or exceeding expectations more if above 100).

Because these scores are relative to where a pupil is expected to be, instead of a raw indicator of the level they are working at, it could be argued that a pupils' reading score staying consistent over time is a neutral outcome - a consistent score implies that a pupil is constantly working at the same level relative to where they should be (i.e. below, at, or above). Therefore, an increase in standardised reading score indicates that a pupil has progressed more than expected, for example they may have been working below expected at baseline and working above expected at endline.

Observing an increase in Bookmark pupils' standardised reading scores is therefore a positive finding as it means that, although still working below the expected level, they are working closer to where they should be than they were previously, and the statistical significance of this change means it can be attributed to the programme. In terms of reading ability, this indicates that their reading ability improved more than anticipated - naturally we expect reading ability to improve over time which is built into how the standardised reading scores are calculated, so a consistent score would reflect the expected amount of progress in a pupil's reading ability.

2.3. Evaluation design

Pupil-level demographic and attainment data was collected from 18 schools. This included data from 265 pupils who were chosen to work with Bookmark (participating) and 648 pupils who were not (control). Attainment data was collected at four timepoints:

Timepoint	When
1	End of Summer term 2023 / Start of Autumn term 2023
2	End of Autumn term 2023
3	End of Spring term 2024
4	End of Summer term 2024

Table 2: Data collection windows

Where a methodology used data collected in the baseline window (Research Questions 1 and 3B), a pupil's baseline data was considered to be the data recorded in the first data collection window, or the second window if this data was not available.

Different research questions used data from different timepoints. All research questions used matched data, whereby all pupils analysed had collected data in all the timepoints of interest.

Attainment data was collected in the form of standardised reading scores, which have a scale of 60-140. The scores are calculated so that age is accounted for, meaning data from pupils of different ages can be compared and aggregated. The scores are also indicators of where a pupil is in relation to a standardised average: if a pupil scores below 100 they are considered below average, and if a pupil scores above 100 they are considered above average.

Research Question 1

This research question aimed to explore the changes made in reading by different subgroups of participating pupils.

The methodology considered **all participating pupils who had at least nearly completed a programme**. This meant that a pupil was considered as 'participating' if they had completed at least five sessions in a single programme at any point throughout the academic year.

Standardised reading scores that were recorded in the baseline, Spring and Summer were used to analyse change over time (see *2.4 Analysis*). Only participating pupils who collected data at all three timepoints were used in the analysis, making a total matched sample size of 191 pupils. **Pupils were separated into different subgroups using the demographic data collected**, so the sample sizes for each subgroup varied, as shown in *Table 3*.

Subgroup		Sample size
All pupils		**191**
Year group	Year 2	14
	Year 3	65
	Year 4	55
	Year 5	57
English as an Additional Language (EAL)	Yes	61
	No	130
Pupil Premium	Yes	63
	No	128
Special Educational Needs (SEN)	Yes	39
	No	150
Gender	Male	103
	Female	88

Table 3: Research Question 1 - sample sizes

Demographic data was also collected for ethnicity, however due to low sample sizes for individual subgroups this data was not analysed fully. Ethnic groups with a sample size of five or more are summarised in *Table 4* – for sample sizes for all ethnic subgroups please refer to *8.1. Appendix – Research Question 2: Ethnicity*. See *2.5. Limitations* and *6.2. Recommendations* for more details). In total, there were 44 ethnicities represented across all participating pupils, as well as ‘N/A’.

Subgroup	Sample size
White British	69
White English	15
Black African	10
Any Other White Background	7
Black Caribbean	7
British	7
Indian	7

Subgroup	Sample size
Asian Pakistani	5
Other Pakistani	5
White and Asian	5
White Cornish	5

Table 4: Research Question 1 – sample sizes for largest ethnic subgroups

Research Question 2

This research question used a control group created by one-to-one propensity score matching (see *2.4 Analysis*) to **compare the outcomes in participating Bookmark pupils with similar pupils at partner schools who were not selected to do a Bookmark programme**. This meant that pupils were 'paired' with each other, and the control and participating groups were of equal size. Pupils were matched on eligibility for Pupil Premium (PP), English as an additional language (EAL) and gender.

The methodology considered participating pupils as pupils who had **at least nearly completed only one programme within a given term (completed five or more sessions), had recorded a reading score at the end of the term before (pre) and at the end of the term in which they participated (post)**. This ensured that the pre and post data was not skewed by any unknown impact of pupils participating in multiple programmes because this could have undermined the reliability and validity of any findings (note that Research Question 3B sought to understand the relationship, if any, between number of programmes participated in and change in reading score to determine whether this unknown impact does indeed need to be accounted for).This method ensured change between pre and post a single programme could be isolated, so that any impact found can be uniquely attributed to the programme rather than as an impact of several programmes. This meant that participating pupils (and their corresponding matched control pupil) may be represented more than once in the sample, but the data used for each window of analysis would be uniquely relative to the programme(s) they engaged with, as outlined in *Table 5*:

Term of participation	Pre data	Post data
Autumn 2023	End of Summer 2023 / start of Autumn 2023	End of Autumn 2023
Spring 2024	End of Autumn 2023	End of Spring 2024
Summer 2024	End of Spring 2024	End of Summer 2024

Table 5: Research Question 2 - timeline of pre/post data

In total, 115 participating pupils were matched with 115 similar control pupils. The nature of Bookmark's delivery model is such that pupil may do multiple programmes, meaning a pupil may have done a programme in each of the three terms. As a result, some pupils were represented in the sample multiple times but with different pre/post data (*Table 5*), meaning the **overall sample size for pre/post data was 144 in both the participating and control groups.** The sample sizes are summarised in *Table 6*:

No. occurrences in data set (terms with one programme only)	Number of pupils (in each participating/ control group)	Overall sample size in dataset (in each participating/ control group)
1	92	92
2	17	34
3	2	6
Total	**111**	**132**

Table 6: Research Question 2 - sample sizes

Research Question 3A

This research question, the first exploring the impact of dosage (the amount of an intervention that a pupil receives) on reading progress, **examined the relationship between the number of sessions within a programme and the change made in reading**. In each programme, pupils could complete up to 12 sessions.

Participating pupils were selected to be in the sample if they **engaged with only one programme within a given term, had recorded a reading score at the end of the term before (pre) and at the end of the term in which they participated (post)**. This meant that data was not influenced by pupils engaging in multiple programmes and the **change between pre and post a single programme could be isolated**. Where pupils completed only one programme in a term more than once (e.g. one programme in Autumn term and one in Spring term), they may be represented in the sample multiple times, but the data used would be relative to each programme they engage with, as summarised in *Table 7*:

Term of participation	Pre data	Post data
Autumn 2023	End of Summer 2023 / start of Autumn 2023	End of Autumn 2023
Spring 2024	End of Autumn 2023	End of Spring 2024
Summer 2024	End of Spring 2024	End of Summer 2024

Table 7: Research Question 3A - timeline of pre/post data

In total, data from 142 pupils across 186 programmes was used. This meant the sample size for pre/post was 186. The sample sizes by number of sessions attended is given in *Table 8*:

Number of sessions completed	Sample size
1	5
2	1
3	2

Number of sessions completed	Sample size
4	8
5	11
6	22
7	19
8	22
9	34
10	28
11	17
12	17
Total	**186**

Table 8: Research Question 3A - sample sizes by number of sessions completed

Research Question 3B

The second of the research questions looking into the impact of dosage on change in reading ability. This research question **investigated the relationship between the number of programmes a pupil did throughout the academic year and change in reading score**.

There were two methodologies for this question, each using a different sample. Firstly, pupils were **categorised by the amount of programmes they had nearly completed (five to seven sessions) or completed (eight sessions or more)**. Participating pupils who collected data at the baseline (end of Summer 2023 / start of Summer 2023) and the end of the Summer 2024 term were used in the sample so that the **change in reading ability could be measured from the start to the end of the 2023/24 academic year**. It should be noted that a programme was considered nearly complete if a pupil did 5-7 sessions, and complete if they did 8 or more sessions. Anything less than this was considered incomplete. Categories are shown in *Table 9:*

Category	Sub-category	Sample size	
Completed none	Incomplete programmes only	5	21
	Nearly completed one only	9	
	Nearly completed multiple, completed none	7	
Completed one	Completed one only	24	47
	Nearly completed one, completed one only	20	

Category	Sub-category	Sample size	
	Nearly completed multiple, completed one	3	
Completed multiple	Nearly completed none, completed multiple	78	130
	Nearly completed one, completed multiple	34	
	Nearly completed multiple, completed multiple	18	
Total		**198**	

Table 9: Research Question 3B - sample sizes by categorisation of pupils by programmes completed

The second approach that this research question took was **exploring the relationship between the change made from the start to the end of the academic year in reading score, and the number of programmes completed** to determine whether the number of programmes completed throughout the year impacted the amount of progress pupils made in reading. Participating pupils were used in analysis if they had recorded a standardised reading score at the baseline (end of Summer 2023 / start of Autumn 2023), at the end of Summer 2024, and had not nearly completed any programmes. This meant that the data was not influenced by pupils partially completing one or more programmes, which helped to understand the impact of pupils completing more programmes, rather than just engaging with more. The sample sizes used are shown in *Table 10:*

Number of programmes completed	Sample size
0	5
1	24
2	36
3	15
4	11
5	12
6	4
Total	**107**

Table 10: Research Question 3B - sample sizes by number of programmes completed

2.4. Analysis

Overview

Research Question	Analysis	Statistical significance tests
RQ1	▶ Measures of central tendency (mean average) ▶ Difference-in-Difference	▶ Independent t-tests ▶ One-sample t-tests ▶ Paired t-tests
RQ2	▶ Measures of central tendency (mean average) ▶ Difference-in-Difference	▶ One-sample t-tests ▶ Paired t-tests
RQ3A	▶ Correlation analysis	▶ T-test for correlation coefficient
RQ3B	▶ Measures of central tendency (mean average) ▶ Correlation analysis	▶ One-sample t-tests ▶ Paired t-tests ▶ T-test for correlation coefficient

Table 11: Analysis and statistical significance testing used by research question

It should be noted that not all analysis has been commented on in this report. Only data with the strongest narratives relating to the research questions and outcomes have been drawn out.

Quantitative data analysis

Various descriptive analysis methods were used to investigate the research questions. For example, **measures of central tendency** were used to calculate the mean average standardised reading scores in various groups of pupils in each relevant data collection. **Comparative analysis** then investigated the difference in the scores in various groups and subgroups across relevant data collection windows. All analysis that involved comparing across multiple timepoints (e.g. pre/post) used data from **matched pupils only**, meaning pupils had to have recorded standardised reading scores in all relevant data collection windows to have been included in the analysis.

Where control groups were used, these were formed by **one-to-one propensity score matching**. This resulted in participating and control groups of equal size. Where propensity score matching was used, both the corresponding participating and control pupil had to have data in all relevant windows, otherwise both pupils were discarded from the analysis.

When looking at the relationship between two variables, for example the number of sessions in a programme and the change made in standardised reading score, the **Pearson correlation coefficient** (r) was used to determine any linear trend. The **coefficient of determination** (R^2) gave insight into the strength of any trends founds.

Frequency distribution analysis was also conducted in places, particularly in quantitative data regarding dosage such as the number of sessions participated in within a single programme. This helped to understand the number of occurrences of each category within the dataset.

Four types of t-tests were used to test data for statistical significance: **paired t-tests, independent t-tests, one sample t-tests and t-tests for correlation coefficients**. All t-tests produced a p-value. The standardised social science convention of a **'significant' p-value being less than 0.05** was used. Where analysis is found to be significant, this suggests that the effect observed in the sample is unlikely to have occurred by chance alone: it rejects the null hypothesis and is instead in favour of the alternative hypothesis.

2.5. Limitations

Some limitations that should be considered when assessing findings include:

- **It was not possible to control the exact amount of Bookmark programmes a pupil participated in.** Bookmark's repeat delivery strategy, where pupils can participate in more than one programme within a term and several throughout the academic year, made it difficult to control for the amount of one-to-one reading the pupil had had. This had various impacts on the analysis methodologies used, such as:
  - **Where engagement in one programme had to be analysed, the window to observe change was only a term long**. This may not have been long enough to observe meaningful change in attainment data.
  - **Pupils in some analyses were aggregated despite engaging in different amounts of the programme**. This may have diluted the effects observed, ignored any relationships between the amount participated in and change in standardised reading scores, and resulted in potential bias as those who engaged more might have differed systematically from those who engaged less.
- **The sample sizes for some subgroups were too small, namely ethnicity**. When teachers shared pupil demographic data, they could use free text to give pupils' ethnicity to account for the inconsistencies in how schools collect this data. However, this meant there was a large range of ethnicities (45) and sample sizes (1-69). There was a larger overlap than anticipated between some groups, but for various reasons – including ethical – these were not grouped together, meaning the majority of sample sizes remained too small to produce meaningful findings.
- **There is a lack of stakeholder voice**. Because this evaluation only measures an outcome using standardised reading scores, there is no pupil, teacher or Bookmark volunteer voice. This means that it gives a narrow view of the programme's impact. Relying solely on attainment data limits the reliability of findings, as pupils only take a test once in each window so could record anomaly scores. The evaluation would benefit from using more measures to assess the outcome.

3. Findings

3.1. Overview of progress made by Bookmark Pupils in 2023/24

5.8

raw increase in standardised reading scores in 2023/24 (n=191, p<0.001)

Pupils who at least nearly completed a Bookmark programme (participated in five or more sessions) recorded an **average increase of 5.8 in their standardised reading scores from the start of 2023/24**, to the end of the Summer term. Furthermore, this was statistically significant (p<0.001). This indicates that pupils' reading ability increased greater than the amount expected (observing no change would have indicated an expected amount of change – see *2.2. Outcome measure: standardised reading scores* for more detail). Nearly 70% of this increase – 4.0 – was made over the Summer term, after pupils had already increased by 1.8 between the start of the academic year and the end of the Spring term. In both periods, the changes observed were statistically significant (p<0.001, p=0.02 respectively), meaning this change was unlikely to have been recorded only on chance.

At the start of the academic year, pupils had an average standardised reading score of 91.2, which was statistically significantly below the benchmark of 100 (p<0.001) as shown in *Figure 1*. However, **the increase seen between the baseline and end of Spring term meant that by the end of Spring term, pupils' average score was 93.0, and by the end of the academic year it had increased further to 97.0**. Though both of these scores were still statistically significantly below the benchmark (p<0.001 both times), it is encouraging to see pupils become a lot closer to it. Should pupils continue to improve their reading scores in the 2024/25 academic year, it is plausible to predict that these Bookmark pupils would become in line with the benchmark, or possibly score above it.

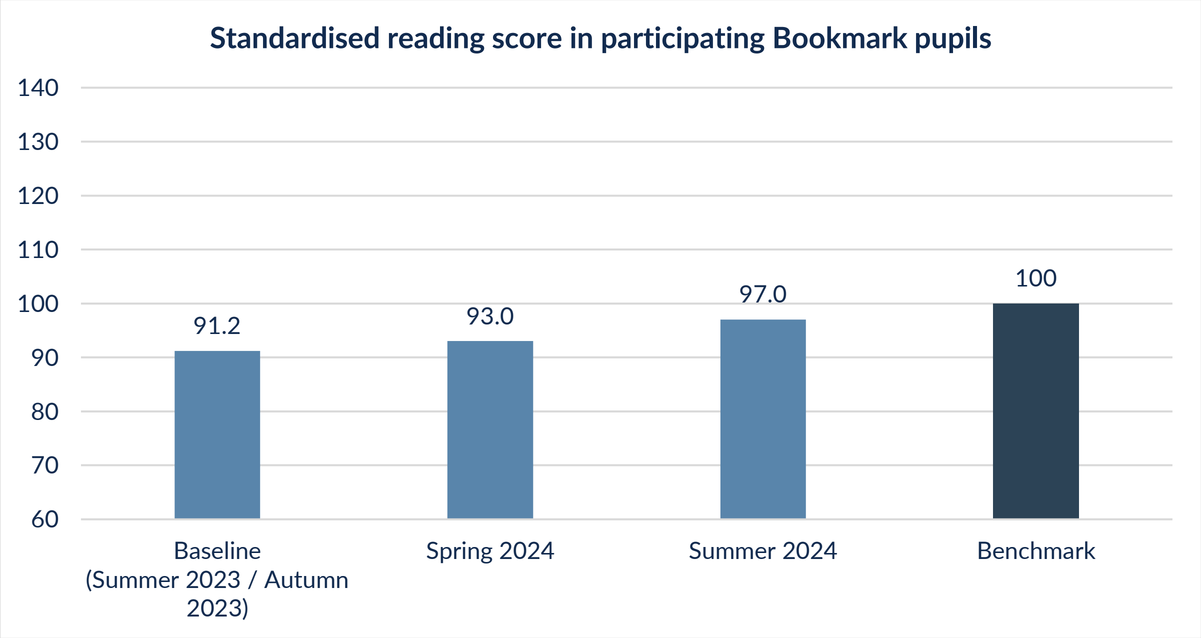


Figure 1: Standardised reading scores in participating Bookmark Pupils in 2023/24 academic year (n=191)

3.2. Research Question 1: Does the Bookmark programme have a differential impact on specific subgroups?

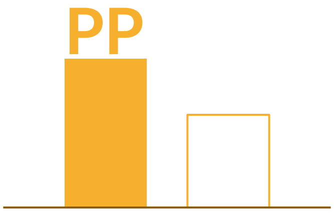
Key findings:

- Pupil Premium pupils' standardised reading scores increased more than in non-Pupil Premium pupils.
- SEN pupils' standardised reading scores increased more than in non-SEN pupils.
- Male and female pupils' standardised reading scores increased by a very similar amount.
- All year groups' standardised reading scores increased. Year 4 pupils' standardised reading scores increased the most out of all the participating year groups. Year 3 pupils increased the least.
- EAL pupils' standardised reading scores did not increase more than non-EAL pupils.
- Of the ethnic groups analysed, all subgroups' standardised reading scores increased.

This section compares the data collected for the following demographic groups:

- Pupil Premium and non-Pupil Premium
- Special Educational Needs (SEN) and non-SEN
- Gender – male and female
- Year group – Years 2, 3 4 and 5
- English as an Additional Language (EAL) and non-EAL

Pupil Premium



...made more progress in reading from the start to end of the 2023/24 academic year than non-PP

Although both pupils eligible for Pupil Premium and pupils who aren't had standardised reading scores below average (100) in all three windows of data collection, **both groups were closer to the average benchmark at the end of the year compared to the start of the year**. Both groups recorded a statistically significant improvement in reading scores from the start to the end of the 2023/24 academic year (p<0.001 for both groups), implying that these increases were not due to chance.

Pupils eligible for Pupil Premium improved the most throughout the year, by 6.7 (p<0.001), while their peers improved by 5.4 (p<0.001). This suggests that the programme may have been particularly beneficial for those who might have more to gain from additional support. Breaking down the academic year, pupils eligible for Pupil Premium recording a higher overall increase is largely due to the changes made from the **start of the year to the end of Spring term, when pupils eligible for Pupil Premium improved by nearly four times as much as non-Pupil Premium pupils**, by 3.5 (p<0.001) and 0.9 (p=0.32) respectively. However, over the Summer term, Pupil Premium pupils' reading scores improved by 3.1 (p <0.001), whereas non-Pupil Premium pupils improved by 4.4 (p<0.001).

However, the differences between the groups' scores at all three windows were not statistically significant (p=0.17 at baseline, p=0.82 in Spring, p=0.63 in Summer), indicating that the differences observed between the two groups was consistently due to chance and unlikely to be as a result of an underlying effect. This means the differences observed in each window should not be considered meaningful. Nevertheless, it was positive to see that **the gap in reading scores between the groups decreased over the academic year**, as illustrated in *Figure 2*. Despite the lack of statistical significance, this does not mean the trend is not potentially meaningful or important. It is recommended than further evaluation is conducted to find a statistical significance between the two groups to confidently determine which one the programme impacted more.

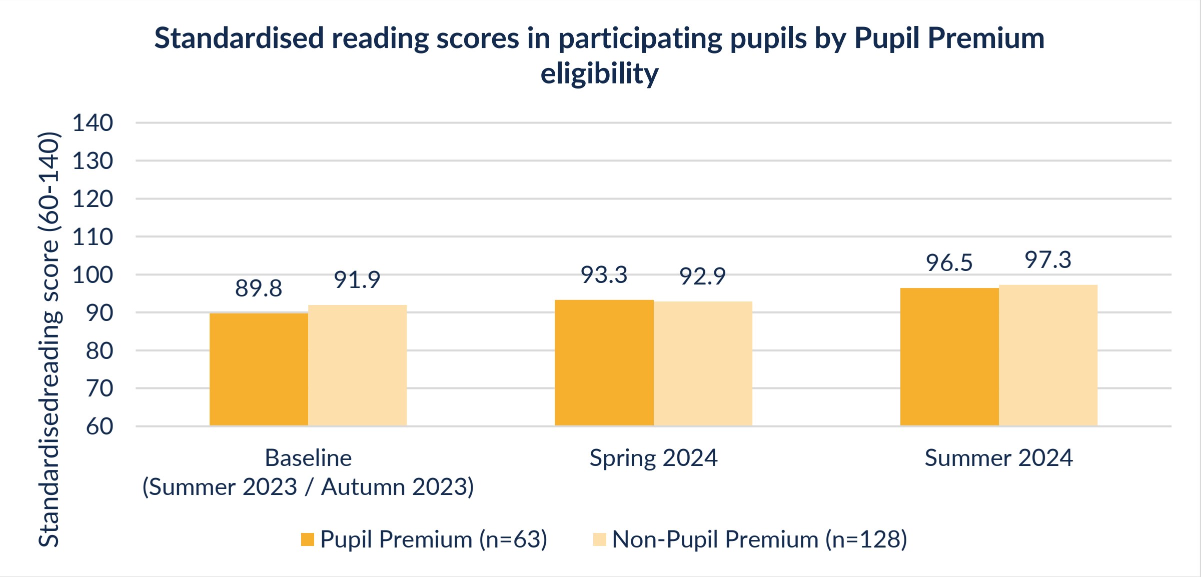
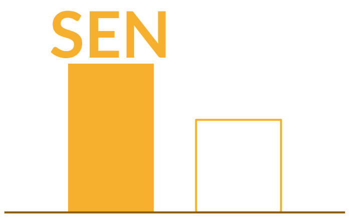


Figure 2: Standardised reading scores in participating pupils in Baseline and Summer (2023/24) by Pupil Premium eligibility

Special Educational Needs (SEN)



...made more progress in reading from the start to end of the 2023/24 academic year than non-SEN

Both pupils with Special Educational Needs (SEN) and pupils without SEN increased their reading scores from the start to the end of the academic year, from 86.1 to 92.8 and 92.7 to 98.1 respectively, as illustrated in *Figure 3*. In all three windows, both groups' reading scores were statistically significant when compared to the national benchmark of 100, which both groups were consistently below (p<0.001 for both groups in all windows). However, **both groups improved in each window**, meaning that by the end of the academic year they were closer to the benchmark than they were at the start of the year.

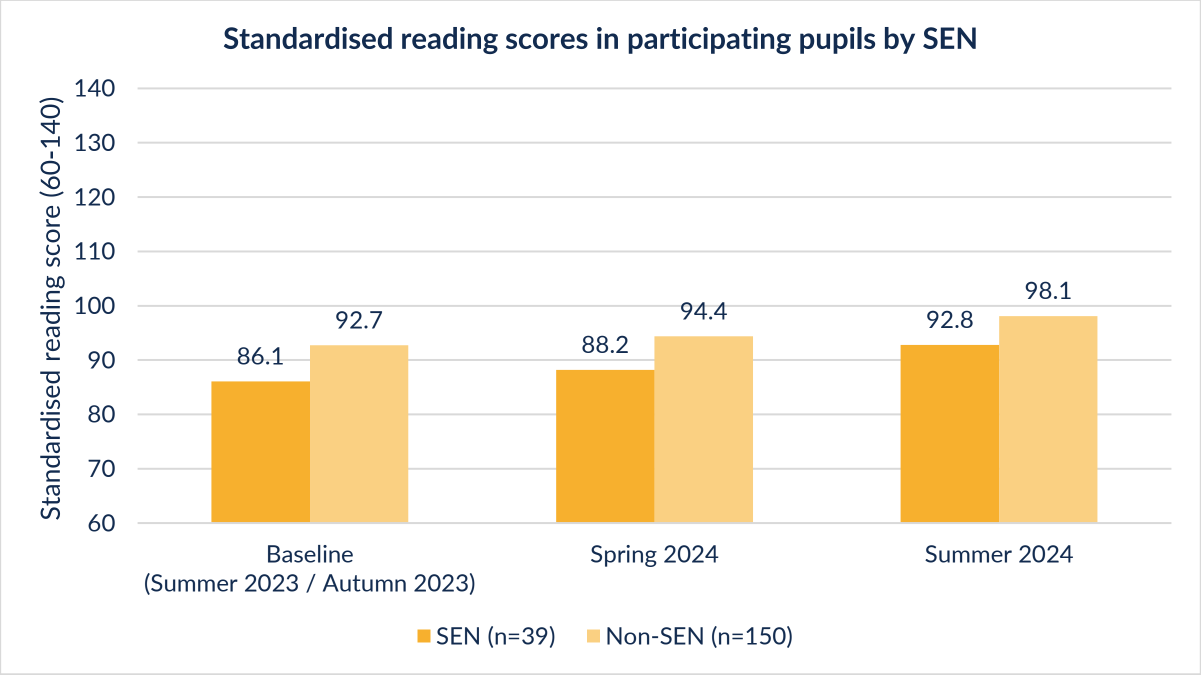


Figure 3: Standardised reading scores in participating pupils in Baseline and Summer (2023/24) by SEN

The **changes in reading scores recorded by both pupils with and without SEN between the start and the end of the academic year were statistically significant** (p<0.001 for both groups). This implies that the improvements made by both groups which resulted in them being closer to the benchmark were unlikely to be due to chance, and possibly a result of the intervention – the use of a control group would be useful to identify whether this was a trend observed in participating pupils, or pupils with and without SEN in general.

From the start to the end of the academic year, **pupils with SEN increased their reading score more than pupils without SEN**, with changes of 6.7 (p<0.001) and 5.4 (p<0.001) respectively. This indicates that Bookmark's programme was effective in targeting pupils who need additional support, particularly as pupils with SEN had consistently lower scores than their peers. As a result, by the end of the year the gap between the two groups had narrowed slightly, from 6.7 at baseline to 5.3 at the end of the year.

Gender

...both made around the same amount of progress from the start to the end of the 2023/24 academic year

Male and female pupils had very similar standardised reading scores at the start and end of the academic year. In the baseline window, male and female pupils had scores significantly below the benchmark of 100 (p<0.001 in both groups), at 91.2 and 91.3 respectively, as shown in *Figure 4*. These scores increased by the end of the Summer term, to scores of 97.0 and 97.1 respectively. Although these scores were still significantly below the average (p=0.01, p=0.02 respectively), it is positive to see pupils get closer to the average. This suggests that **the programme did not target a specific gender** and have a greater impact on one than the other over the course of the academic year.

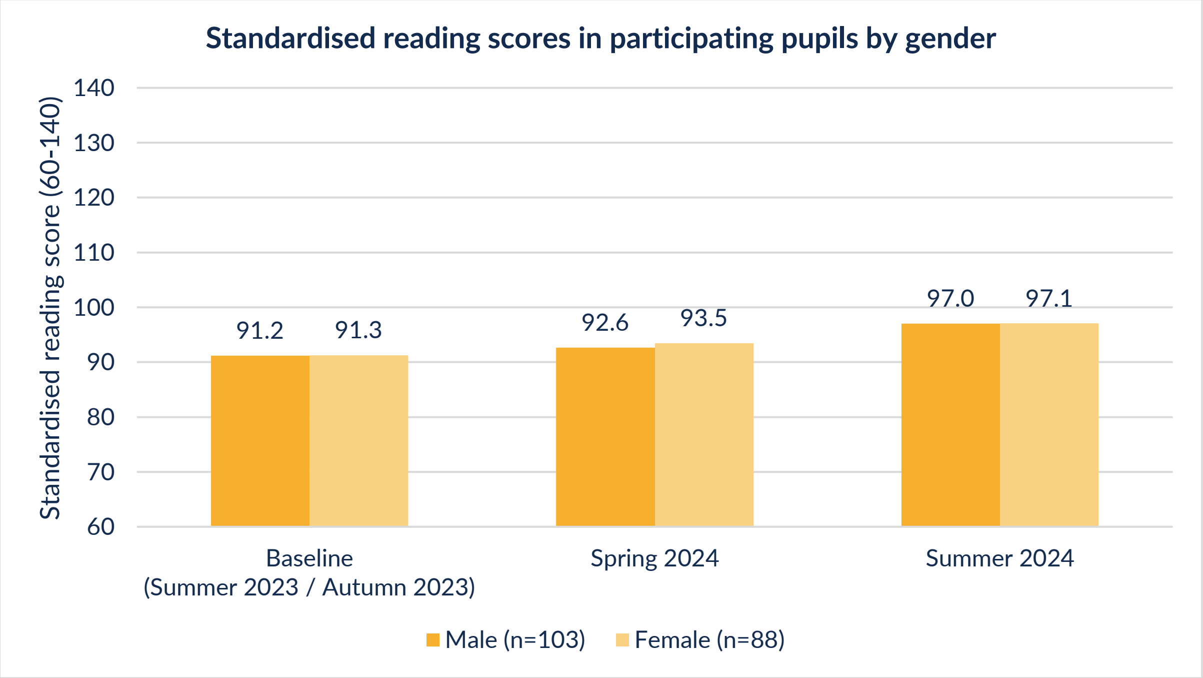
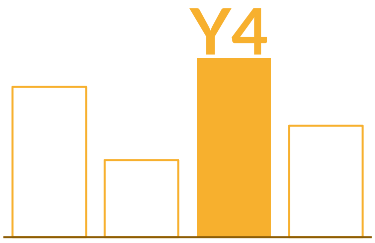


Figure 4: Standardised reading scores in participating pupils in Baseline and Summer (2023/24) by gender

Furthermore, the **overall changes in reading scores in both groups from the start to the end of the academic year were statistically significant** (p<0.001 for both), implying that these improvements are not due to chance and could be the impact of the programme over the academic year.

The only difference between the two groups' reading scores was observed when **female pupils scored slightly higher than male pupils at the end of the Spring term**, with scores of 93.5 and 92.6 respectively. However, the difference between the groups was not statistically significant (p=0.63).

Year group



...made the most progress in reading from the start to end of the 2023/24 academic year

Pupils in years 1 to 5 participated in the Bookmark Reading programme, however data was only collected for pupils in years 2 to 5, so only these year groups were analysed in this evaluation. Standardised reading scores were recorded for all these year groups. Because these scores account for the age of pupils, they can be used as a measure of reading that is comparable across year groups.

As shown in *Figure 5*, pupils in **all year groups improved from the start of the 2023/24 academic year to the end.** In all cases, this change was statistically significant (p=0.01 for Year 2, p<0.001 for all other year groups), indicating that this change was unlikely to be due to chance. **Pupils in Year 4 made the most progress during this period, improving from an average standardised reading score of 92.9 to 101.4. This made them the only year group to score above average (100) in any of the terms throughout the year**, although the difference between Year 4 pupils' scores and the average was not statistically significant (p=0.32), so could be due to chance instead of as a result of the intervention.

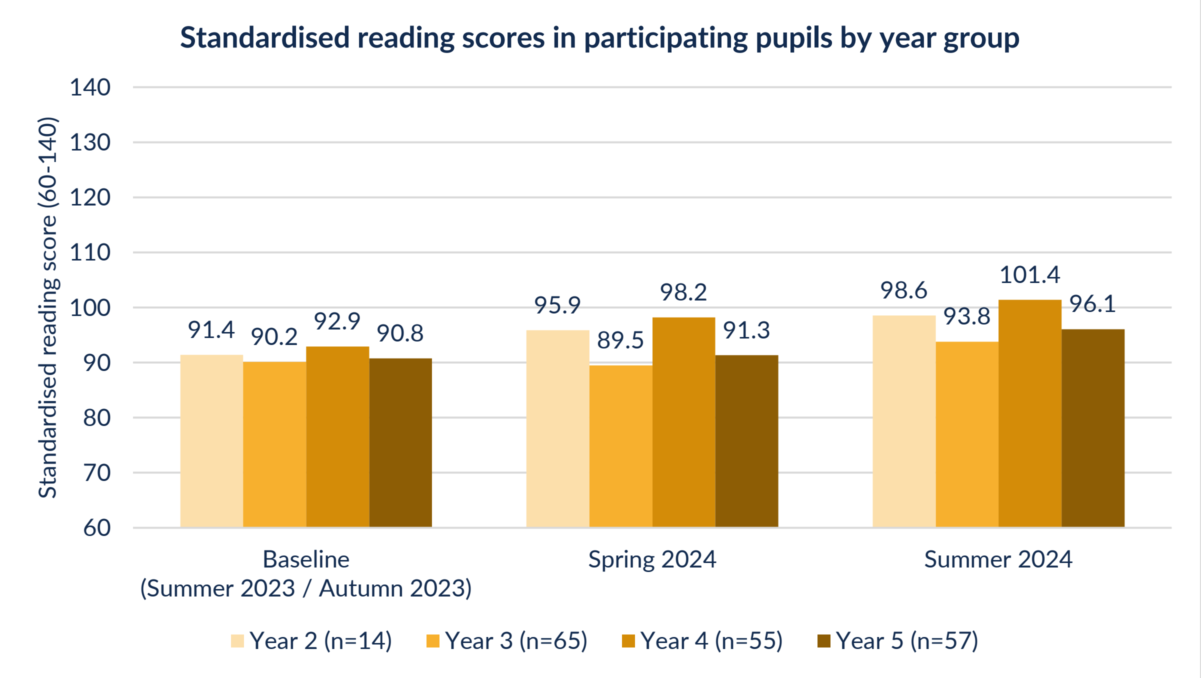


Figure 5: Standardised reading scores in participating pupils in Baseline and Summer (2023/24) by year group

Year 3 pupils were the only year group who showed any sign of a decreased reading score throughout the year, though they did finish the year with a better reading score than what they started with. They had a baseline reading score of 90.2 – the lowest of any year group – which decreased slightly to 89.5 in the Spring term, although this change was not significant (p=0.59). It should be noted that the nature of standardised reading scores means that a reduced score does not necessarily mean a regression in reading ability, but does indicate that pupils are falling further below average for pupils of their age. Nevertheless, they improved significantly (p<0.001) between the Spring and Summer terms, making their final reading score in the academic year 93.8. This was still the lowest score of all the year groups, but it was positive to observe an overall, statistically significant change from the start of the year (p<0.001).

As illustrated in *Figure 6*, **Year 5 pupils had a similar trend to Year 3 pupils**, where they made better improvement on their standardised reading scores between Spring and Summer, when they improved by 4.7, than between the start of the academic year and the end of Spring term, when they improved only slightly by 0.6.

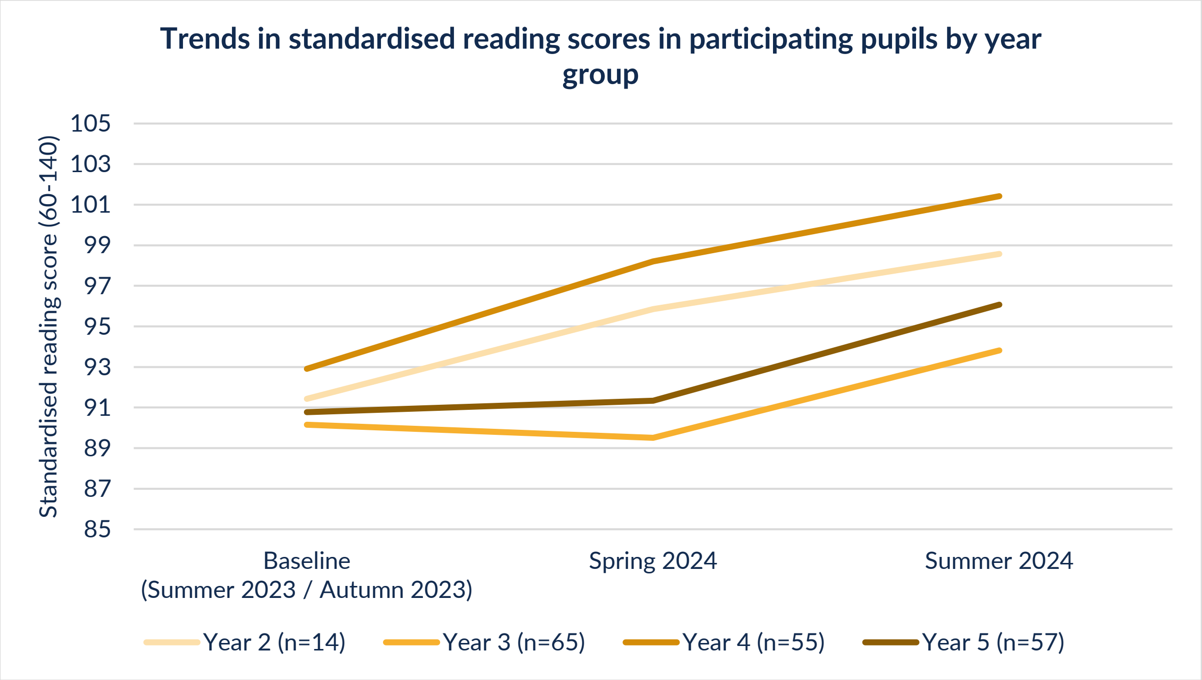


Figure 6: Trends in standardised reading scores in participating pupils in 2023/24 by year group

Contrastingly, **both Year 2 and Year 4 recorded a greater improvement in their reading score between the baseline and end of Spring term than they did over the Summer term.** Year 2 pupils' reading scores significantly increased by 4.4 (p=0.04) between the baseline and end of Spring term, then increased further still by 2.7 over the Summer term (p=0.24). Similarly, Year 4 pupils' reading scores initially improved by 5.3 (p<0.001), and then further still by 3.2 (p<0.001), both changes of which were statistically significant.

This made Year 4 pupils the only year group where both periods of change were statistically significant, indicating that their short-term improvement during the year was unlikely to be due to chance. Nevertheless, **all year groups' change in reading scores from the start to the end of the year was statistically significant** (p=0.01 for Year 2, p<0.001 for all other year groups). This suggests that Bookmark's programme appears to have had a positive effect on reading scores for all year groups overall throughout the year, although it would be useful to use a control group to understand whether this was instead just a general trend observed, and not attributed to Bookmark's programmes itself.

English as an Additional Language (EAL)

...made less progress in reading from the start to end of the 2023/24 academic year than non-EAL

Both pupils with English as an Additional Language (EAL) and non-EAL pupils increased their reading scores over the academic year, from 92.4 to 97.3 and 90.7 to 96.9 respectively, as shown in *Figure 7*. This meant that **while non-EAL pupils made slightly more progress than EAL pupils, EAL pupils continued to score very slightly higher than their peers**. This difference was very small; the gap between the two groups narrowed to only 0.4 by the end of the year. Scores for both groups were also in line with each other in the Spring term, when EAL pupils scored 93.1 and non-EAL pupils scored 93.0. Furthermore, the differences between the groups' scores in all three windows were not statistically significant (p=0.28 in baseline, p=0.95 in Spring, p=0.84 in Summer). This indicates that there was no meaningful difference between the groups throughout the year.

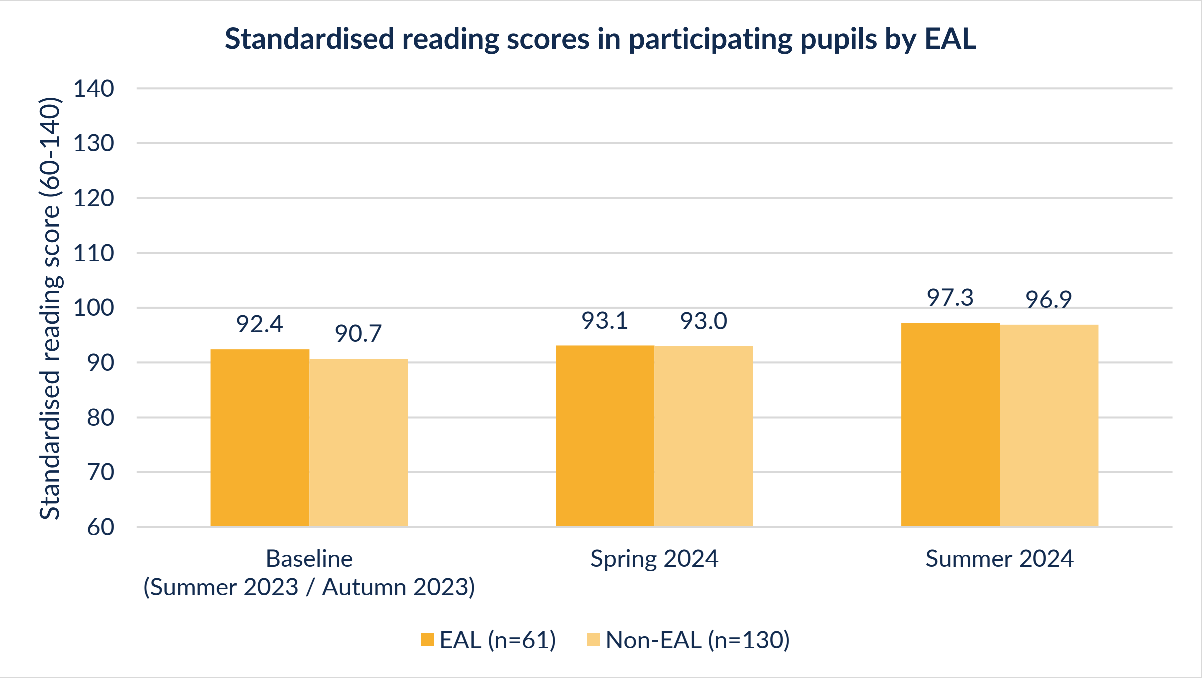

Standardised reading scores in participating pupils by EAL

Figure 7: Standardised reading scores in participating pupils in Baseline and Summer (2023/24) by EAL

Ethnicity

As noted in *2.3. Evaluation design*, not all ethnic subgroups were analysed. Due to the small sample sizes for individual subgroups, no statistical analysis was conducted in this section. Therefore, findings should be interpreted with caution – they are useful as indicative findings but require further analysis and evaluation to warrant as meaningful. Where analysis for a particular subgroup has been discussed, the sample size (n) has also been given.

Of the ethnic groups with a sample size of five or more (11 groups in total), **all subgroups recorded an increase in standardised reading scores between the start and end of the academic year**, as shown in *Table 12*. **'White Cornish' pupils increased the most** with a change of 15.0 (n=5), while **'Indian' pupils made the least progress** with a change of 3.3 (n=7). It's worth noting that 'White Cornish' pupils had the lowest reading score at the start of the year, at 83.4, but their progress meant they were more in line with the benchmark (100) at 98.4 at the end of the year. This meant that at the end of the year, they had the sixth highest reading score of all 11 groups.

Subgroup	Standardised reading score			Change in standardised reading score		
	Baseline	Spring 2024	Summer 2024	Baseline-Spring	Spring-Summer	Baseline-Summer
White British	91.4	93.0	96.0	+1.64	+2.97	+4.6
White English	88.3	92.6	98.7	+4.33	+6.07	+10.4
Black African	91.8	90.5	97.1	-1.30	+6.60	+5.3
Any Other White Background	89.9	96.9	99.6	+7.00	+2.71	+9.7
Black Caribbean	86.6	89.6	96.4	+3.00	+6.86	+9.9
British	95.4	98.9	104.0	+3.43	+5.14	+8.6
Indian	94.9	98.4	98.1	+3.57	-0.29	+3.3
Asian Pakistani	93.4	96.4	101.0	+3.00	+4.60	+7.6
Other Pakistani	89.0	91.6	93.8	+2.60	+2.20	+4.8
White and Asian	98.2	102.6	102.8	+4.40	+0.20	+4.6
White Cornish	83.4	96.4	98.4	+13.00	+2.00	+15.0

Table 12: Research Question 1 - ethnic subgroup analysis

At the start of the year, no groups were above the national average of 100 – the highest scoring subgroup was 'White and Asian', at 98.2. By the end of the Spring term, this was the only subgroup to have a reading score above the benchmark, at 102.6. **In the Summer term, three groups scored above the benchmark**: 'British' at 104.0 (n=7); 'White and Asian' at 102.8 (n=5); 'Asian Pakistani' at 101.0 (n=5).

3.3. Research Question 2: Do Bookmark pupils make more progress in reading than similar non-Bookmark pupils?

Key findings:

- Bookmark pupils had a lower baseline standardised reading score than similar non-Bookmark pupils, indicating the pupils most in need of help with their reading were selected to participate in Bookmark's programme.
- 41% of participating pupils made more progress than a similar (matched) pupil.
- Further analysis using more data is needed to meaningfully answer Research Question 2. In this analysis, the difference between the amounts of progress made by both groups was small and not statistically significant.

Pupils who had participated in a single Bookmark programme during any given term were matched with similar pupils who did not do a Bookmark programme at all, based on eligibility for Pupil Premium, English as an additional language (EAL) and gender. Both groups' standardised reading scores were collected before and after the participating pupil took part in a programme. Note that this is a different methodology to the sections *3.1.* and *3.2.* (overview of progress made by Bookmark pupils throughout 2023/24 and findings for Research Question 1).

Before the programme, control pupils on average had a higher reading score than Bookmark pupils, by 10.5, as shown in *Figure 8*. Furthermore, participating pupils' reading scores were statistically significantly below the national average of 100 by 6.3 (p<0.001), while control pupils were statistically significantly above the national average by 4.3 (p<0.001). This suggests that **the pupils most in need of help with their reading were selected to participate in Bookmark's programme.**

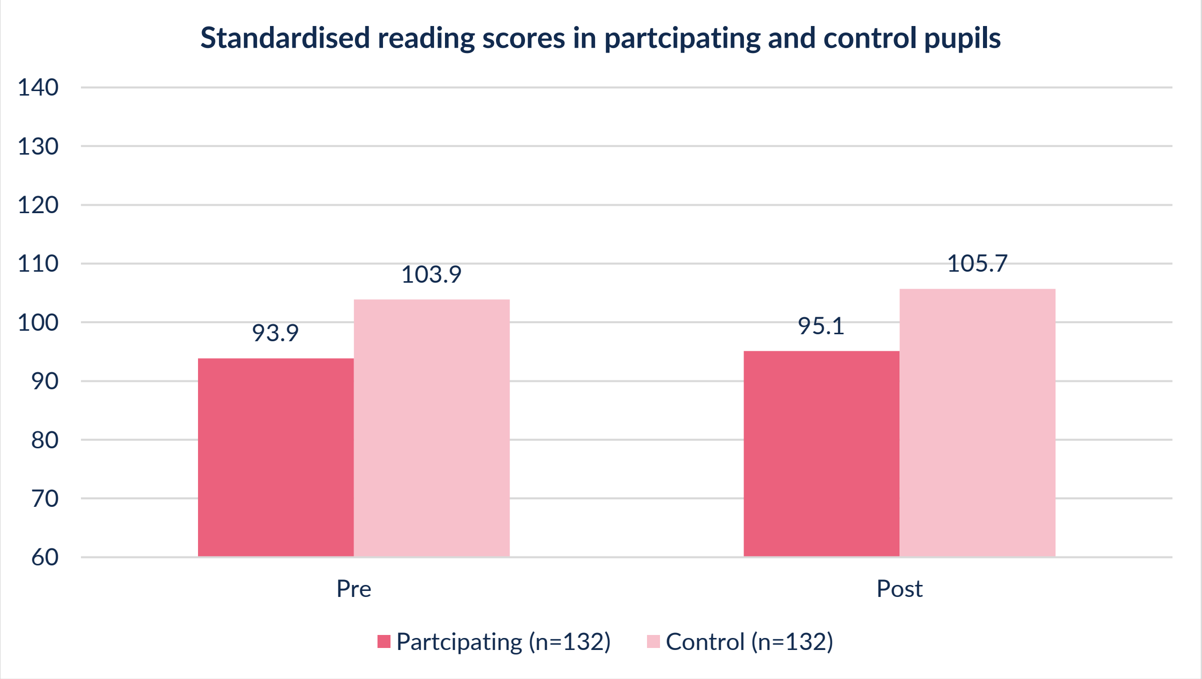
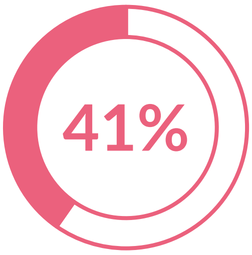


Figure 8: Standardised reading scores in similar participating and control pupils before and after a programme

However, this difference in baseline scores also highlights that the participating and control group were not similar in reading ability, presenting a limitation to this methodology. It was not possible to match pupils based on this because the control sample was taken from current partner Bookmark schools, meaning pupils with lower standardised reading scores were identified as in need of the programme and consequently selected to participate in it, leaving only pupils with higher reading scores available to be in the control group. This made it difficult to fully answer this research question.

Both the participating and control groups recorded increases in their reading scores after the participating pupils did a Bookmark programme, control pupils made a slightly greater improvement, of 1.8, which was statistically significant (p=0.04). Participating pupils, meanwhile, improved by 1.2, but this was not statistically significant (p=0.10). While this indicates that Bookmark pupils do not, in the short-term at least, on average make more progress in reading than similar non-Bookmark pupils after doing one programme, it should be noted that both increases were very small and similar, and **the difference between the change observed in both groups was not statistically significant** (p=0.61). This difference is therefore nothing more than a suggestive finding as it could be due to chance, so it would be beneficial to explore this further to find a statistically significant difference, which would allow for a more definitive conclusion to the research question.

As a result of the changes observed, **participating pupils remained statistically significantly below the national average (p<0.001) while control pupils were statistically significantly above it** (p<0.001). It would be interesting to investigate if this finding is still evident when looking at pupils who have done multiple programmes over a longer period of time. Nevertheless, it was positive to observe participating pupils becoming closer to the national average.



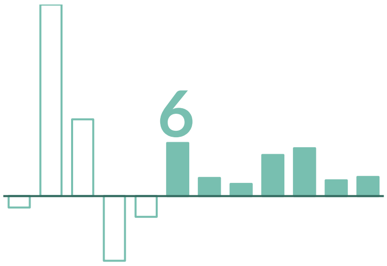
...of participating pupils made more progress than a similar (matched) control pupil

Furthermore, through brief explorative analysis of this dataset it was found that **41% of participating Bookmark pupils made more progress than the non-Bookmark pupil they were matched with**. It would be beneficial to delve into this finding more to understand if there are any common attributes or characteristics in this 41%, such as gender or Pupil Premium status.

3.4. Research Question 3A: Does the number of sessions in a programme impact the amount of reading progress a pupil makes?

Key findings:

- There was no clear linear correlation between the number of sessions in a programme and the amount that a pupil's standardised reading score changed.
- Pupils who had participated in 6-12 sessions recorded a consistent average increase in standardised reading score, while those who complete 1-5 sessions in a single programme recorded average changes that were less consistent.



Pupils who had participated in 6-12 sessions recorded a consistent average increse in standardised reading score

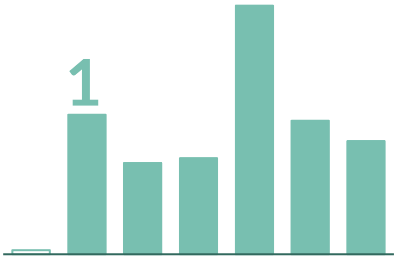
Pupils who had completed up to five sessions within a single programme recorded average changes in standardised reading scores that fluctuated heavily, with pupils who attended one, four and five sessions of a programme recording decreases of up to 3.4. However, **pupils who attended 6-12 sessions of a programme recorded a consistent average increase in standardised reading score**. This indicates that pupils should **participate in at least six sessions, as it suggests that they could be more confident that they would improve their reading score**. Although, it should be noted that no changes were statistically significant. This meant that none of the changes recorded by the grouped pupils can be confidently attributed as an impact of the programme because they are likely to have been made by chance. Yet, these insignificances are likely to be due to small sample sizes (n=1 to 34).

However, there was **no clear linear correlation between the number of sessions a pupil attended within a single programme and the change they recorded in their standardised reading score**, as reflected in a Pearson's R score of 0.05 (R^2=0.00, p=0.52) and illustrated in *Figure 8*. This considered the change a pupil made from the term before participating in the programme, to the end of the term that they participated in. This meant that although pupils were more consistently above average if they completed six or more sessions, pupils did not necessarily have a tendency to increase their reading score more if they did more sessions.

3.5. Research Question 3B: Does the number of programmes a pupil participates in impact the amount of reading progress they make?

Key findings:

- Pupils who completed at least one programme increased their standardised reading score by a greater amount than those who did not complete any.
- There was a very weak positive correlation between the number of programmes pupils completed throughout the academic year and the change in their standardised reading scores.
- Pupils who at least nearly completed a programme recorded a greater increase in reading age than those who had not.



Pupils who completed one or more programmes improved their reading scores more than those who didn't complete any.

The nature of Bookmark's delivery strategy means that pupils were able to complete multiple programmes within the academic year. Pupils were said to have completed a programme if they participated in eight or more sessions of a single programme.

When analysing the average amount of change pupils made based on the number of programmes they completed, it is evident that **those who completed at least one increased their reading scores by a greater amount than those who did not complete any**, as shown in the infographic to the left. Pupils who completed no programmes throughout the year only improved by 0.2 (p=0.97), meaning their reading score remained very steady. However, pupils who completed any number of programmes improved by at least 4.4, with pupils who had participated in four programmes recording the largest increase, by 12.1 (p=0.00). This could indicate that four is the appropriate number of programmes for a pupil to complete for them in order to improve the most they can within a year, particularly as the change observed was statistically significant, although the low sample sizes should be considered (only 11 pupils completed four programmes) – this analysis was not part of the original research plan, so the data collection methods were not specifically tailored for it. However, these indicative findings are promising and should be explored further in future evaluations.

Furthermore, there was a **positive linear correlation, albeit very weak and not statistically significant, between the number of programmes pupils completed throughout the academic year, and the change in their standardised reading scores** (r=0.11, R^2=0.01, p=0.27). This implies that there is a slight tendency for pupils, who participated in more programmes to show a greater improvement in their reading scores, as shown in *Figure 9* (on the following page). However, this relationship is extremely weak and indicates that the number of programmes a pupil completes is not a strong predictor of changes in reading scores.

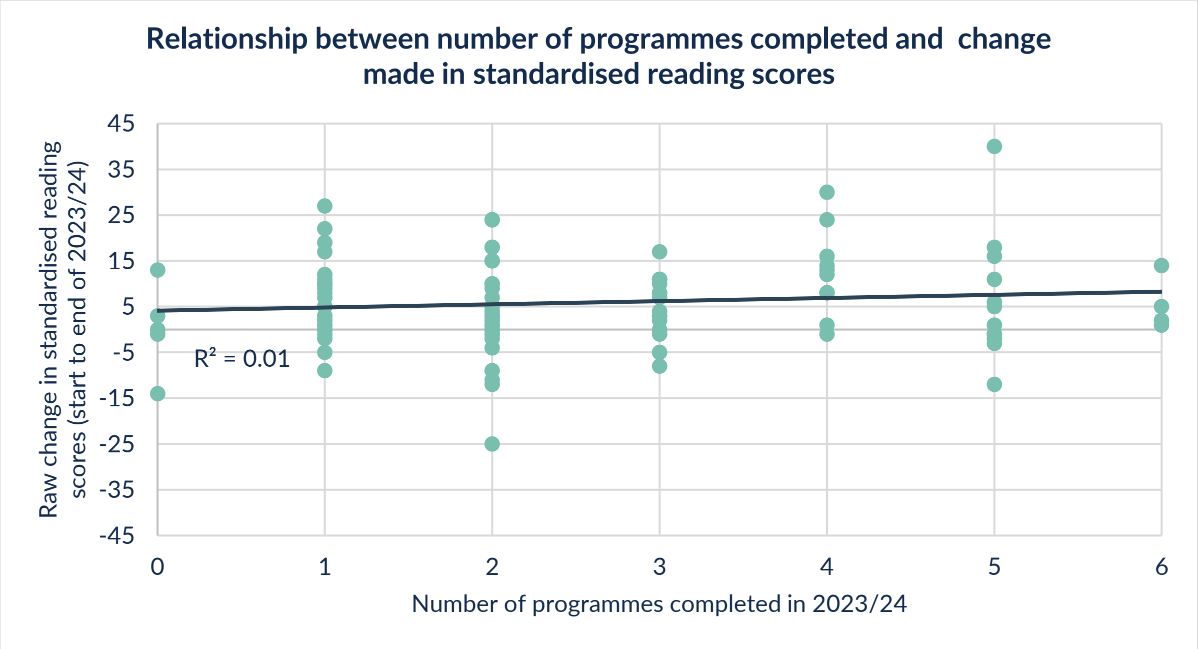


Figure 10: Relationship between number of programmes in 2023/24 and change made in standardised reading scores, from the start to the end of the academic year

Some pupils participated in Bookmark's programmes without completing them. Pupils were noted as nearly completing a programme if they participated in five to seven sessions. Programmes where pupils participated in four or less sessions were considered incomplete. To understand the impact of dosage further, participating pupils were categorised by the number of sessions and programmes they did, as shown in *Figure 10.*

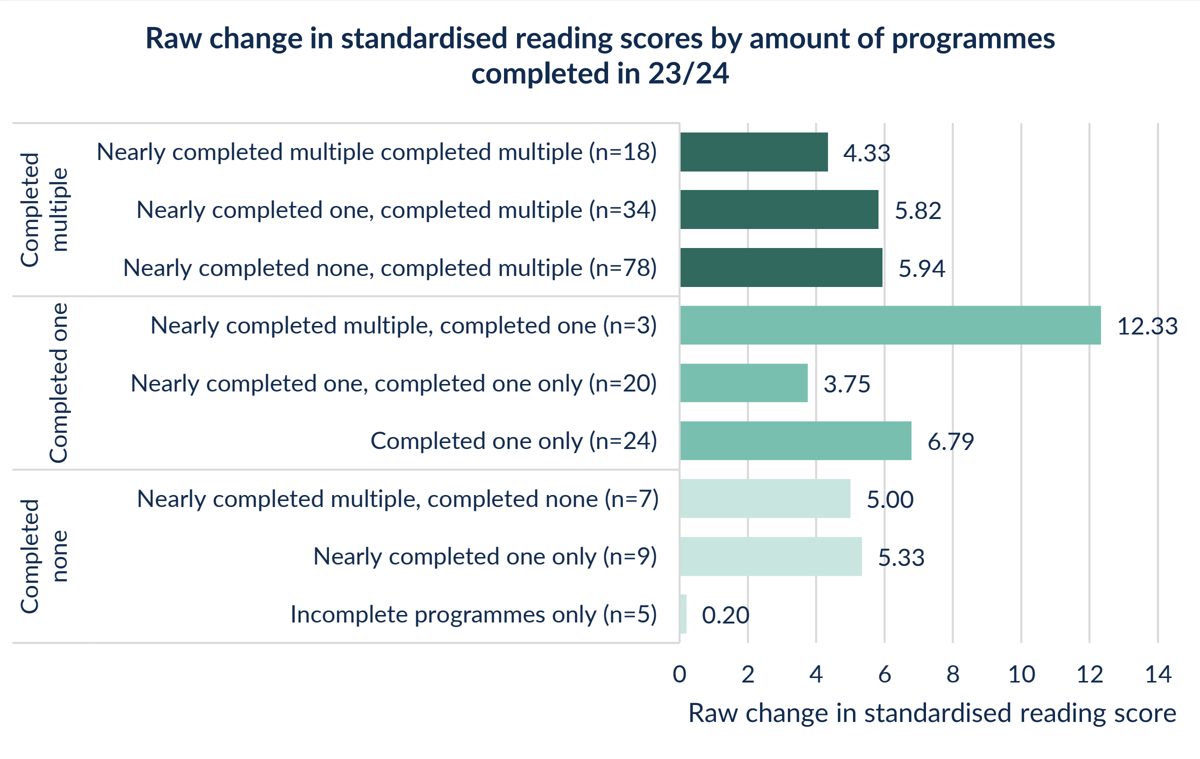


Figure 9: Change in standardised reading scores by amount of programmes completed, from the start to the end of the academic year

It was again clear that **pupils who did not complete any programmes, nor nearly complete any programmes ('Incomplete programmes only'), recorded a lesser improvement in their reading score** than those who at least nearly completed a programme. Pupils who did not complete or nearly complete any programmes only progressed by 0.2 (p=0.97), while pupils who at least nearly participated in one improved by at least 3.8.

However, in the pupils that did at least nearly complete a programme, there was **no clear trend in the relationship between the number of programmes and sessions completed, and the change pupils made in their reading scores**. Those who nearly completed multiple programmes as well as completing one recorded the biggest increase, of 12.3, although this was not statistically significant (p=0.12) and the small sample size should be noted (n=3).

6. Conclusion

This evaluation looked to investigate the impact of Bookmark Reading Charity's programme on academic progress in pupils in years 2-5. Firstly, it aimed to understand whether the programme helped Bookmark pupils make more progress in reading than similar non-Bookmark pupils. Secondly, it investigated whether this impact was even across all subgroups, or whether it supported some pupils in improving their reading better than it supported others. Lastly, it sought to understand the impact of dosage, both through the number of sessions per programme that pupils attended as well as the number of programmes pupils completed throughout the academic year.

6.1. Summary of findings

Outlined below are noteworthy findings for each research question, as well as an overview summary of the progress made by Bookmark pupils throughout the 2023/24 academic year.

Progress made by Bookmark pupils throughout 2023/24

- Pupils who at least nearly completed a Bookmark programme (participated in five or more sessions) recorded an **average increase of 5.8 in their standardised reading scores from the start of the 2023/24** (p<0.001).
- On average, pupils had a **baseline standardised reading score of 91.2**. This increased to **93.0 by the end of the Spring term**, and **97.0 by the end of the Summer term**.

1 Does the Bookmark programmes have a differential impact on specific groups?

- **Pupil Premium pupils' standardised reading scores increased more than in non-Pupil Premium pupils**. Pupil Premium pupils' scores increased by 6.68, whereas non-Pupil Premium pupils' increased by 5.4 over the academic year. This meant that the **gap between the two groups' reading scores had narrowed from the start of the year**, to only 0.8.
- **SEN pupils' standardised reading scores increased more than in non-SEN pupils**. SEN pupils increased their score by 6.7 while non-SEN pupils' scores increased by 5.4. Both changes were statistically significant (p<0.001 for both). This indicates that **Bookmark's programme was effective in providing targeted support for SEN pupils**.
- **Male and female pupils' standardised reading scores increased by the same amount**, by 5.8. Both changes were also statistically significant (p<0.001 for both). This meant **the programme did not favour supporting a particular gende**r, indicating gender equality in impact.
- **Pupils in each of the four year groups recorded an overall increase in standardised reading score. Year 4 pupils' standardised reading scores increased the most out of all the participating year groups; Year 3 pupils increased the least**. Both of these changes were statistically significant (p<0.001 for both). **Year 4 pupils were the only subgroup across the evaluation to exceed the national benchmark** reading score of 100, with a

score of 101.4 in the Summer term. However, the difference between this score and the benchmark was not statistically significant (p=0.3).

- **EAL pupils' standardised reading scores did not increase as much as in non-EAL pupils.** EAL pupils' scores increased by 4.9, while non-EAL pupils' scores increased by 6.2. Both changes where statistically significant (p<0.001 for both). However, this closed the gap between the two groups scores as **EAL pupils had a slightly higher score at both the start and end of the academic year**, by 1.7 and 0.4 respectively.

2 Do Bookmark pupils make more progress in reading than similar non-Bookmark pupils?

- **Pupils most in need of help with their reading were selected to participate in Bookmark's Programme.** This was evident in control pupils, on average, having a higher standardised reading score than participating pupils, by 10.1.
- **41% of Bookmark pupils made more progress than a similar (matched) non-Bookmark pupils.** This implies that there are some Bookmark pupils making more progress than similar non-Bookmark pupils. It is recommended that this finding is investigated further to understand if there are particular factors contributing to this trend (e.g. are they mostly Pupil Premium pupils?).
- **Further analysis is required to conclusively answer this research question**, as the difference between the amounts of progress made by both groups was found to be small and not statistically significant (p=0.61). Nevertheless, Bookmark and non-Bookmark pupils recorded an increase in reading scores from before to after a programme (+1.2, p=0.10 and +1.8, p=0.04 respectively).

3A Does the number of sessions in a programme impact the amount of reading progress a pupil makes?

- There was **no clear linear correlation between the number of sessions in a programme and the amount that a pupil's standardised reading score changed** (r=0.05, R^2=0.00, p=0.52). This suggests that further investigation into the impact of dosage and the number of sessions pupils attended would be beneficial to understand how the number of sessions contributes to the positive change in readings scores observed in research questions one and two.
- **Pupils who had participated in 6-12 sessions in a programme recorded a consistent average increase in standardised reading score**, while those who completed 1-5 sessions in a single programme recorded average changes that were less consistent, with those participating in one, four and five sessions recording decreases. This suggests that **those who attend at least six sessions can be more confident to see an increase in their standardised reading score** than a decrease, although in line with the finding above, the more sessions a pupil completed does not necessarily suggest that they will improve more in reading.

3B Does the number of programmes a pupil participates in impact the amount of reading progress they make?

- **Those who completed at least one programme increased their reading scores by a greater amount than those who did not complete any.** Pupils who completed no programmes throughout the year only improved by 0.2 (p=0.97), meaning their reading score remained very steady. However, pupils who completed any number of programmes improved by at least 4.4. Although, the low sample sizes and varying degrees of statistical significance here should be considered and further evaluation is recommended.
- Similar to the above finding, **pupils who at least nearly completed a programme recorded a greater increase in reading age than those who had not**. Although, there was no further clear trend observed between the number of programmes pupils had nearly completed or completed, and the change recorded in their standardised reading scores. Again, further evaluation is recommended to conclusively answer this research question.
- There was a **very weak positive linear correlation between the number of programmes pupils completed throughout the academic year and the change in their standardised reading scores** (r=0.11, R^2=0.01, p=0.27). Although this is likely due to chance, this could serve as an indicator that **those who complete more programmes have a slight tendency to record a greater increase in the standardised reading scores**.

6.2. Recommendations

Recommendations for any future iterations of the programme include:

- **Encourage pupils to complete a minimum of six sessions**. When investigating Research Question 3A, it was found that those who completed 6-12 sessions recorded a consistent average increase in standardised reading score, while those who completed 1-5 sessions in a single programme recorded average changes that in some cases decreased. Therefore, it is recommended that pupils complete at least six sessions as it is suggested that they would be more likely to increase their standardised reading score, rather than decrease it.
- **Continue to not limit the number of programmes a pupil can do over the year**. As found when exploring Research Question 3B, there is a weak positive correlation between the number of programmes completed, and the improvement recorded in their standardised reading score. Therefore, pupils should be encouraged to do more programmes.
- **Consider tailoring the programme for different year groups more.** Findings from Research Question 1 showed that while all year groups' standardised reading scores increased, the amount by which scores changed varied, resulting in the range of scores increasing. This indicates that the programme in its current form targets some year groups better than others, so it could be worth adjusting the design of it to address these disparities.
- **Continue targeting Pupil Premium pupils and those with Special Educational Needs**. As found in Research Question 1, Pupil Premium pupils made more progress in reading over year than their peers. Similarly, those with SEN also increased their reading score more than non-SEN pupils. This suggests that the intervention has been successful in targeting these pupils, an impact that the programme should aim to sustain.

Recommendations for any future evaluation regarding the project include:

- **Review the methodology for Research Questions 2 and 3A to allow a longer window to observe impact.** Currently, these research questions only allow windows to be one term long, so that data was collected just before and after a programme and that data would not be influenced by the complication of pupils participating in multiple programmes throughout the year. However, this is a small period of time to observe impact in attainment data. Consider a methodology where longer-term data is collected for more pupils who only engage in or complete one programme throughout the year, so that longer term impact – if any – can be observed and robustly analysed.
- **Conduct subgroup analysis with participating and control pupils.** This report has determined there is differential impact on specific pupils in Research Question 1. The inclusion of a control group would enhance the validity and reliability of these findings. It would also enable further investigation into Research Question 2 to understand why some Bookmark pupils made better progress in reading than similar non-Bookmark pupils, but on overall average non-Bookmark pupils' scores were increasing slightly more.
- **Run a retrospective evaluation that matches participating pupils with similar pupils from newly partnered Bookmark schools, using baseline standardised reading scores as covariates in propensity score matching.** Propensity score matching was used in

Research Question 2, but only used demographic data. This data helps match pupils based on similar backgrounds but does not consider pupils' current reading ability and how this compares to the national benchmark. For example, a pupil who is already notably above the benchmark might find it harder to increase their score compared to a pupil notably below it. Matching on baseline score would enable comparison between pupils who are academically similar. It's worth noting that this could not be done in this evaluation due to pupils with lower reading scores being selected to participate in the programme. Therefore, doing a retrospective evaluation using pupils from newly partnered Bookmark schools who, if the school had partnered the year previously, would have done a Bookmark programme would enable matching to be done based on similar reading abilities. Furthermore, this would allow further exploration into the finding from Research Question 3B regarding pupils who completed at least one programme improving more than pupils who did not complete any but were meant to.

- **Re-investigate Research Question 2 (the amount Bookmark pupils progress compared to none-Bookmark pupils) using a methodology that looks at progress over the entire academic year**. The methodology used in this evaluation used pre/post data collected when a pupil had only done one programme to control for the amount of intervention a pupil had received. However, findings in Research Question 3B suggested there was no statistically significant linear correlation between the number of programmes a pupil participates in and the change made in reading score, implying that this is not something that needs to be controlled for. Instead, looking at change over an entire year rather than term would give more time for change in reading score to occur, making statistical findings more likely.
- **Analyse and interpret stakeholder voice alongside attainment data.** This evaluation only uses snapshot attainment data which limits the reliability of findings. Bookmark already collect qualitative data from teachers and volunteers through their ongoing evaluation cycle and are planning on collecting qualitative data from pupils too. It would be beneficial to analyse and interpret this data alongside attainment data, as a single evaluation, to build a holistic understanding of the impact of the programme.
- **Ask teachers to select an appropriate ethnic group for pupils, with the added option of 'Other (please state)',** rather than allowing them to state their ethnicity through free text. The variation and overlap of ethnic groups submitted by teachers via open text, though somewhat anticipated, was underestimated for this evaluation which made it challenging to analyse subgroup data fully. Doing this would mean the number of ethnic subgroups is more limited, similar groups can be aggregated together and, as a result, sample sizes are larger. This would allow for analysis between different ethnic groups to understand whether the programme has a differential impact.

7. Glossary

7.1. Evaluation terminology

Academic attainment

This refers to test scores in academic subjects such as maths, science, English etc. Some evaluations will compare pupils' attainment in tests for these subjects at the start (baseline) and end (final) of an evaluation to see whether they have made progress over time.

Baseline

The initial assessment of pupils' attainment or social and emotional skills, at the start of an evaluation.

Change over time

The difference between a pupil's baseline result and their final result, either for attainment or social and emotional skills. This indicates progress made during participation in the programme. This will begin to indicate whether the programme has had an impact on pupils, though we must also account for other factors that could lead to this change, which is why we recommend the use of control groups and qualitative analysis.

Control Group

A control group is composed of students who do not participate in the programme and who closely resemble the pupils who take part in the programme in attainment and demographic traits. It is used to get an indication of whether a change in results over the course of the programme can likely be attributable to the programme itself, or whether results were likely to change over time in any case. Also known as a comparison group.

Evaluation

An evaluation is set up to measure the impact of a particular programme. This will involve monitoring the programme over a specified period, for one or more groups, in order to evaluate the progress participating pupils make. One programme can involve multiple evaluations, and we recommend gathering data across multiple time points to ensure valid and reliable results are generated.

Evaluation Group(s)

An evaluation will either cover one specific group of pupils, who all participate in the programme (e.g. a new programme trialled in one class, or an intervention with one small group). Or, the evaluation may cover multiple evaluation groups (e.g. as several small-group interventions, or with multiple classes carrying out the same programme). In the case of multiple evaluation groups, it can be useful to compare the outcomes for different groups to build up a stronger data set, as well as to compare differences in implementation to see whether this has an effect on results.

Final

The final assessment of pupils' attainment or social and emotional skills at the end of an evaluation.

Matched Pupils

Matched pupils are pupils who carried out a baseline, midpoint and a final assessment at the start, during and end of the evaluation. It can be useful to consider results from matched pupils because this means only including those pupils who participated in the full duration of the programme.

Outcomes

We use outcomes to refer collectively to any social and emotional skills and academic attainment scores that are being measured over the course of an evaluation.

Participating pupils

The group of pupils participating in the evaluation, and not forming part of a control group.

Programme

This could be any intervention, project or programme run in school with the aim of improving pupil outcomes or life chances.

7.2. Statistical analysis terminology

Independent t-test

Use to compare scores between two different groups. This can be used when comparing data recorded by a participating group to a benchmark.

One sample t-test

Used to compare the mean of a single sample to a known or hypothesised population mean.

One-to-one propensity score matching

This is a statistical technique used to create a comparison between two groups (participating and control) that are similar according to the recorded demographic data. This reduces selection bias. The use of a control group using propensity score matching improves the validity of the evaluation by controlling for confounding variables, ensuring the findings are more meaningful.

Statistically significant

A result has statistical significance when it is very unlikely to have occurred given the null hypothesis. In other words, if a result is statistically significant, it is unlikely to have occurred due purely to chance.

P Value

A p-value is a measure of the probability that an observed result could have occurred by chance alone. The lower the p-value, the greater the statistical significance of the observed difference. Typically a p-value of ≤ 0.05 indicates that the change was statistically significant. A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates strong evidence for the null hypothesis, i.e. that we cannot be confident that this change did not occur due purely to chance.

Paired t-test

Used to see if the difference between pre/post scores in a matched group is statistically significant.

7.3. Education terminology

English as an Additional Language

Pupils with English as an Additional Language (EAL) refers to learners whose first language is not English.

Pupil Premium (PP)

The pupil premium grant is designed to allow schools to help disadvantaged pupils by improving their progress and the exam results they achieve. Whether a child is eligible for Pupil Premium funding is often used by schools as an indicator of disadvantage.

Special Educational Needs (SEN) (sometimes referred to as SEND)

A child or young person has special educational needs and disabilities if they have a learning difficulty and/or a disability that means they need special health and education support; this is usually shortened to SEND.

8. Appendix

8.1. Research Question 2: Ethnicity

Sample sizes

These are pupils who have at least nearly finished a Bookmark programme and had Standardised reading scores that were recorded in the baseline, Spring and Summer.

Subgroup	Sample size
White British	69
White English	15
Black African	10
Any Other White Background	7
Black Caribbean	7
British	7
Indian	7
Asian Pakistani	5
Other Pakistani	5
White and Asian	5
White Cornish	5
Any Other Mixed Background	4
Pakistani	4
White (Eastern European)	4
Any Other Black Background	3
ARAB	2
Asian Indian	2
Black British	2
Turkish	2
Albanian	1
Asian	1
Asian and Any Other Ethnic Group	1
Bangladeshi	1
Black Nigerian	1
Black Somali	1
Gypsy / Roma	1

Subgroup	Sample size
Iranian	1
Kashmiri Pakistani	1
Kurdish	1
Mexican	1
Mirpuri Pakistani	1
N/A	1
Other White British	1
Pashto	1
Polish	1
ROM	1
Romanian	1
Sri Lankan Tamil	1
Vietnamese	1
White (Polish)	1
White and Black	1
White and Black African	1
White and Black Caribbean	1
White European	1
White other	1

www.evaluation.impactedgroup.uk

Supporting our purpose driven partners to make better decisions using high quality evidence.

Get in touch

hello@impacted.org.uk

www.evaluation.impactedgroup.uk

ImpactEd Evaluation

ImpactEd Group

Limited Company number 14266872